Probability and Statistics

Chapter 6

SAMPLING DISTRIBUTIONS AND ESTIMATION

Dr. Ali Wagdy

UST-Fall 2023

Contents

1. Basic Definitions.
2. Sampling Distribution.
3. Sampling Distribution of $\overline{X}$.
4. Proportions.
5. Point Estimation.
6. Interval Estimation.
7. Confidence Interval For The Proportion P.

Basic Definitions

- **Population:**
  - The largest collection of the observations with which we are concerned.
  - The largest collection of values of a random variable for which we have an interest at a particular time; Population may be finite or infinite.
- **Population Distribution:**
  - The probability distribution of population data.

In many cases it is not possible to obtain information about all members of a population, for the following reasons:

1. The collecting of the information may destroy the population.
2. The population may be infinite.
3. Even if a measurement could be made for each member of a population, considerations of time and expense usually dictate otherwise.

Basic Definitions

- **Sample:**
  - A subset of population data.
  - If a sample of size **n** is drawn from a population in which a way that every possible sample of size **n** has the same probability of being selected, the sample is called a **random sample**.

- **Statistic:**

is any quantity whose value can be calculated from sample data. There is uncertainty as to what value of any particular statistic will result. Therefore, a statistic is a random variable and will be denoted by an uppercase letter.

Sampling Distribution

- **Introduction:**

The value of a **population parameter** is always constant. For example, for any population data set, there is only one value of the **population mean** $(\mu)$and **Variance** (σ^2)**.** However, we cannot say the same about the **sample mean** $(\bar{X})$, We would expect different samples of the same size drawn from the same population to yield different values of $\bar{X}$ which is considered as a **random variable** that possesses a probability distribution, which is more commonly called **sampling distribution of** $\bar{X}$ **.**

Other sample statistics such as the sample variance S^2 also possesses sampling distributions.

Sampling Distribution

- **Sampling distribution:**

The probability distribution of all possible values which can be assumed by some **statistic**, computed from samples of the same size randomly drawn from the same population, is called the ***sampling distribution*** of that statistic.

The ***sampling distribution*** of a statistic will depend on:

1. The population size.
2. The samples' size.
3. The method of choosing the random samples.

Sampling Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from a population with mean μ and variance σ^2. The **sample mean** and the **sample variance** are:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{And} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Where $\bar{X}$ and S^2 are considered as random variables.

- The mean of the variable $\bar{X}$ is always equal to the mean of the population from which the random samples are chosen and doesn't depend on the size of the sample.

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

Sampling Distribution

- The variance of $\bar{X}$ depends on the sample size and is equal to the original population variance σ^2 divided by **n**.

$$\mathbf{var}(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

- The mean of the variable of S^2 is equal to the original population variance σ^2.

$$E(S^2) = \sigma^2$$

- The square root of the variance of $\bar{X}$ is called the **standard error of the mean**. Consequently, the larger the sample size, the smaller the standard error of $\bar{X}$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Sampling Distribution

Example 1:

A certain brand of MP3 player comes in three configurations: a model with 2 GB of memory, costing \$80, a 4 GB model priced at \$100, and an 8 GB version with a price tag of \$120. If 20% of all purchasers choose the 2 GB model, 30% choose the 4 GB model, and 50% choose the 8 GB model, then the probability distribution of the cost X of a single randomly selected MP3 player purchase is given by.

X	80	100	120
$P(X)$	0.2	0.3	0.5

$$\mu = 106\,;\, \sigma^2 = 244$$

Sampling Distribution

Suppose on a particular day only two MP3 players are sold. Let X_1 the revenue from the first sale and X_2 the revenue from the second. Suppose that X_1 and X_2 are independent, each with the probability distribution shown before; then:

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
80	80	.04	80	0
80	100	.06	90	200
80	120	.10	100	800
100	80	.06	90	200
100	100	.09	100	0
100	120	.15	110	200
120	80	.10	100	800
120	100	.15	110	200
120	120	.25	120	0

$\bar{x}$	80	90	100	110	120
$p_{\bar{X}}(\bar{x})$	.04	.12	.29	.30	.25

s^2	0	200	800
$pS^2(s^2)$	.38	.42	.20

Sampling Distribution

$$\mu_{\bar{X}} = E(\bar{X}) = \sum \bar{x} p_{\bar{X}}(\bar{x}) = (80)(.04) + \cdots + (120)(.25) = 106 = \mu$$

$$\sigma^2_{\bar{X}} = V(\bar{X}) = \sum \bar{x}^2 \cdot p_{\bar{X}}(\bar{x}) - \mu^2_{\bar{X}}$$

$$= (80^2)(.04) + \cdots + (120^2)(.25) - (106)^2$$

$$= 122 = \frac{244}{2} = \frac{\sigma^2}{2}$$

$$E(S^2) = 0.42 * 200 + 0.2 * 800 = 244 = \sigma^2$$

- Sampling distribution of $\bar{X}$ is centered at the population mean μ and the sampling distribution of S^2 is centered at the population variance σ^2.
- Sampling distribution of $\bar{X}$ becomes more concentrated about μ as the sample size n increases.

Sampling Distribution of $\bar{X}$

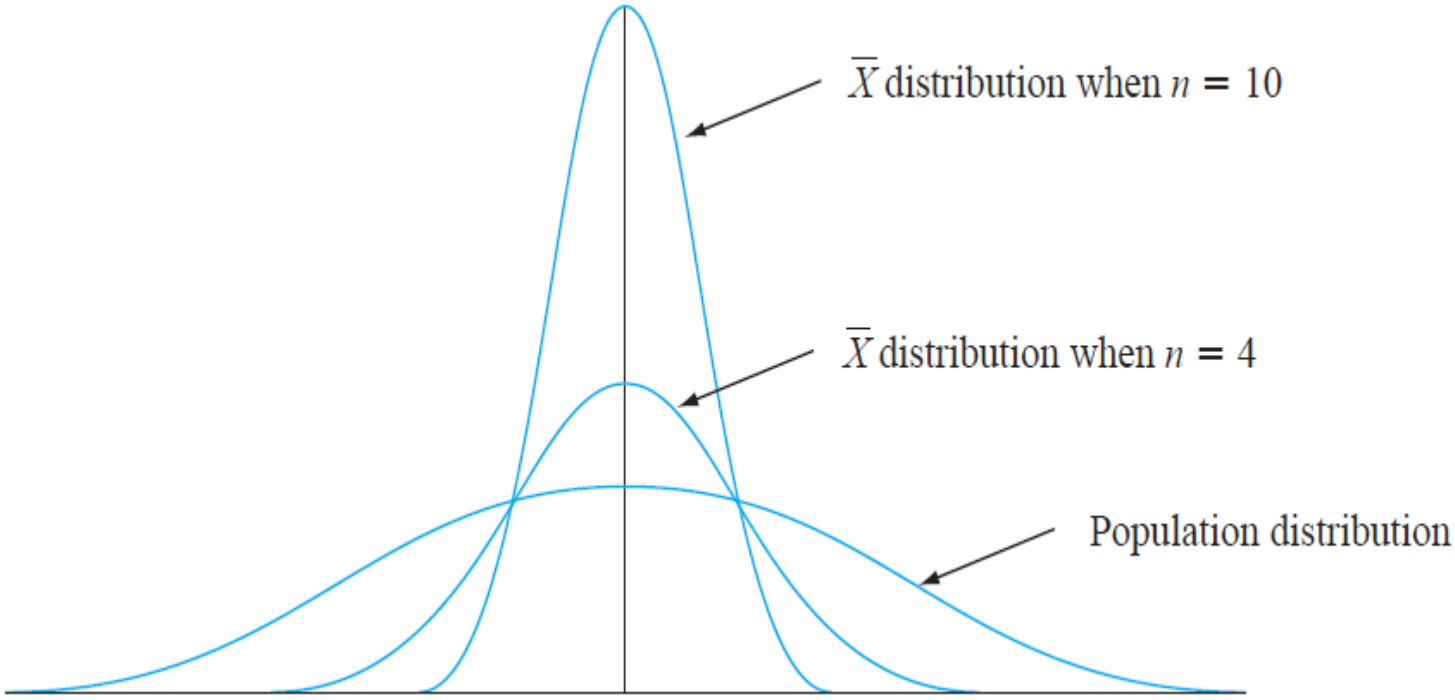
- The importance of the sample mean $\bar{X}$ springs from its use in drawing conclusions about the population mean μ, Some of the most frequently used inferential procedures are based on properties of the sampling distribution of $\bar{X}$.

- **Case (1) (the population is normal with known σ^2, or n is large):**

When the population from which samples are drawn is normally distributed with mean μ and **known** variance σ^2 , then the sampling distribution of $\bar{X}$ will be normal

with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$; i.e. $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

Sampling Distribution of $\overline{X}$



Sampling Distribution

Example 2:

The time that it takes a randomly selected rat of a certain subspecies to find its way through a maze is a normally distributed **rv** with $\boldsymbol{\mu = 1.5\ min}$. and $\boldsymbol{\sigma = 0.35\ min}$. Suppose five rats are selected. Let $\boldsymbol{X_1, X_2, \dots, X_5}$ denote their times in the maze. what is the probability that the sample average time is at most 2.0 min?

Solution:

$\boldsymbol{\mu_{\bar{X}} = \mu = 1.5}$, and $\boldsymbol{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.35}{\sqrt{5}} = 0.1565}$

$$\boldsymbol{P(\bar{X} \le 2) = P\left(Z \le \frac{2 - 1.5}{0.1565}\right) = P(Z \le 3.19) = F(3.19) = 0.9993}$$

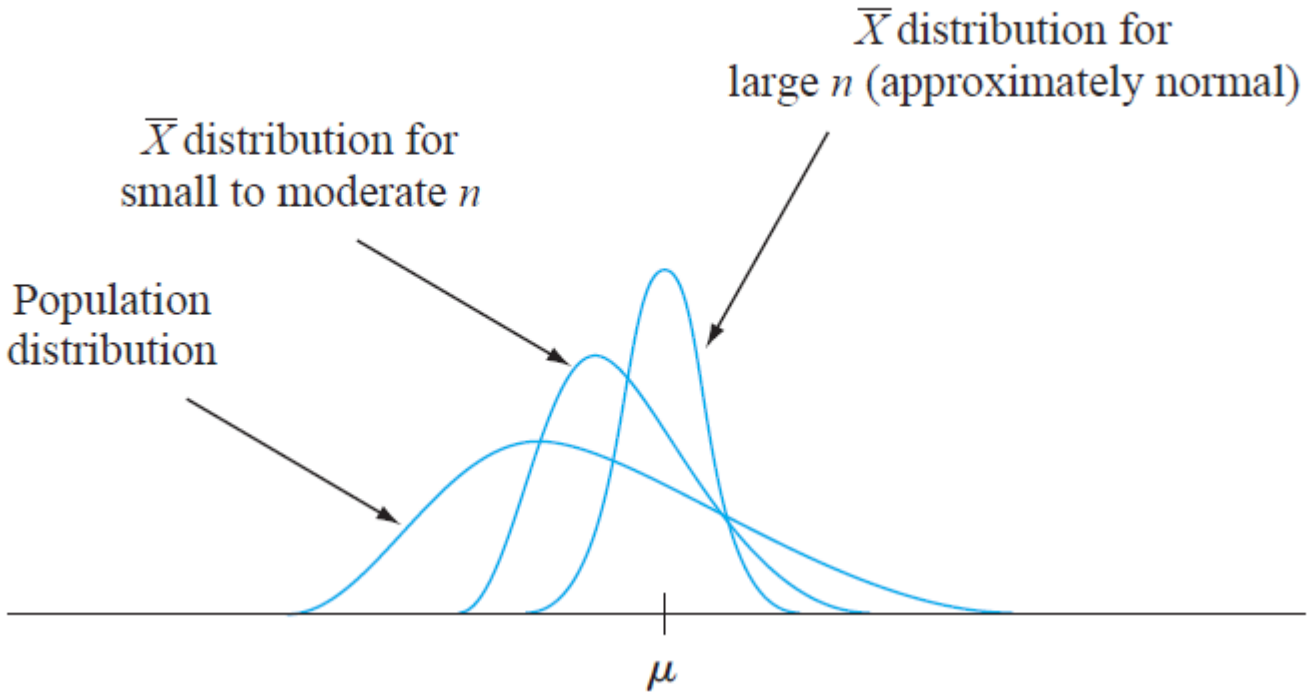
Sampling Distribution of $\bar{X}$

- **Central Limit Theorem:**

➢ If random samples of **large enough size n** are drawn from any population (**non-normal**) with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma^2}$, the sampling distribution of $\bar{X}$ will be approximately normal with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$; i.e. $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

➢ The Normal approximation in the central limit theorem will be good if **n ≥ 30** irrespective of the shape of the population distribution. If the population variance $\boldsymbol{\sigma^2}$ is unknown, the central limit theorem still valid when we replace $\boldsymbol{\sigma^2}$ by the sample variance $\boldsymbol{S^2}$, i.e. for large **n** enough, we have $Z = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim N(0,1)$

Sampling Distribution of $\overline{X}$



Sampling Distribution of $\overline{X}$

Example 3:

The amount of a particular impurity in a batch of a certain chemical product is a random variable with mean value 4.0 g and standard deviation 1.5 g. If 50 batches are independently prepared, what is the (approximate) probability that the sample average amount of impurity $\overline{X}$ is between 3.5 and 3.8 g?

Solution:

$$n = 50\,,\ \mu_{\overline{X}} = \mu = 4, \text{ and } \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{50}} = 0.2121$$

$$P(3.5 \le \overline{X} \le 3.8) = P\left(\frac{3.5-4}{0.2121} \le Z \le \frac{3.8-4}{0.2121}\right) = P(-2.357 \le Z \le -0.943)$$

$$= F(2.357) - F(0.943) = 0.164$$

Sampling Distribution of $\bar{X}$

- **Case (2) (n is small and the population is normal with unknown σ^2):**

According to central limit theorem; the sampling distribution of $\bar{X}$ can be approximated by standard normal distribution even for unknown population variance σ^2 but this is valid for large sample size **(n ≥ 30).** What if the population variance σ^2 is unknown and sample size less than **30** for normal distribution population!!?

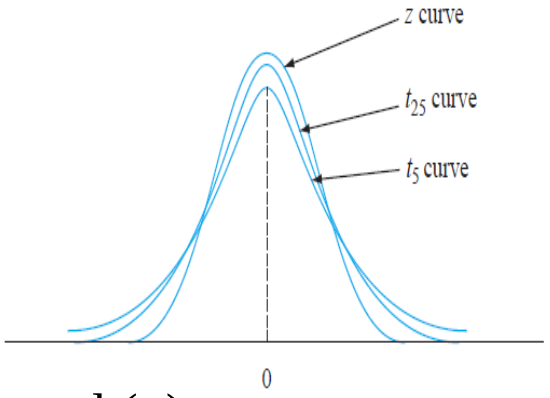
Theorem 1:

If $\bar{X}$ and S^2 are the mean and variance, respectively, of a random sample of size **n** taken from a population that is **normally distributed** with mean μ and unknown variance σ^2. Then the random variable $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ has a t-distribution with $\nu = n - 1$ degrees of freedom (df).

Sampling Distribution of $\bar{X}$

- The normal distribution is governed by two parameters (μ, σ); But for t distribution it results from specifying the value of a single parameter ($\boldsymbol{\nu}$), called the **number of degrees of freedom,** abbreviated df.

- **Properties of t Distributions:**

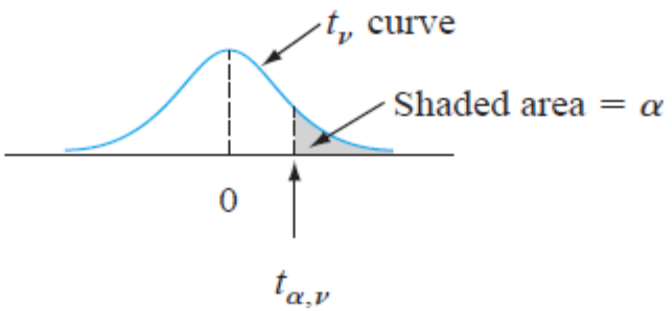


Let $\boldsymbol{t_\nu}$ denote the $\boldsymbol{t}$ **distribution** with $\boldsymbol{\nu}$ df.

1. Each $\boldsymbol{t_\nu}$ curve is bell-shaped and centered at **0**.
2. Each $\boldsymbol{t_\nu}$ curve is more spread out than the standard normal ($\boldsymbol{z}$) curve.
3. As $\boldsymbol{\nu}$ increases, the spread of the corresponding $\boldsymbol{t_\nu}$ curve decreases.
4. As $\boldsymbol{\nu} \to \infty$ the sequence of $\boldsymbol{t_\nu}$ curves approaches the standard normal curve (so the $\boldsymbol{z}$ curve is often called the t curve with $df = \infty$).

Sampling Distribution of $\bar{X}$

- **Notation:**



- Let $t_{\alpha,\nu}$ = the number on the measurement axis for which the area under the *t* curve with ν df to the right of t is α (called a **t critical value**).

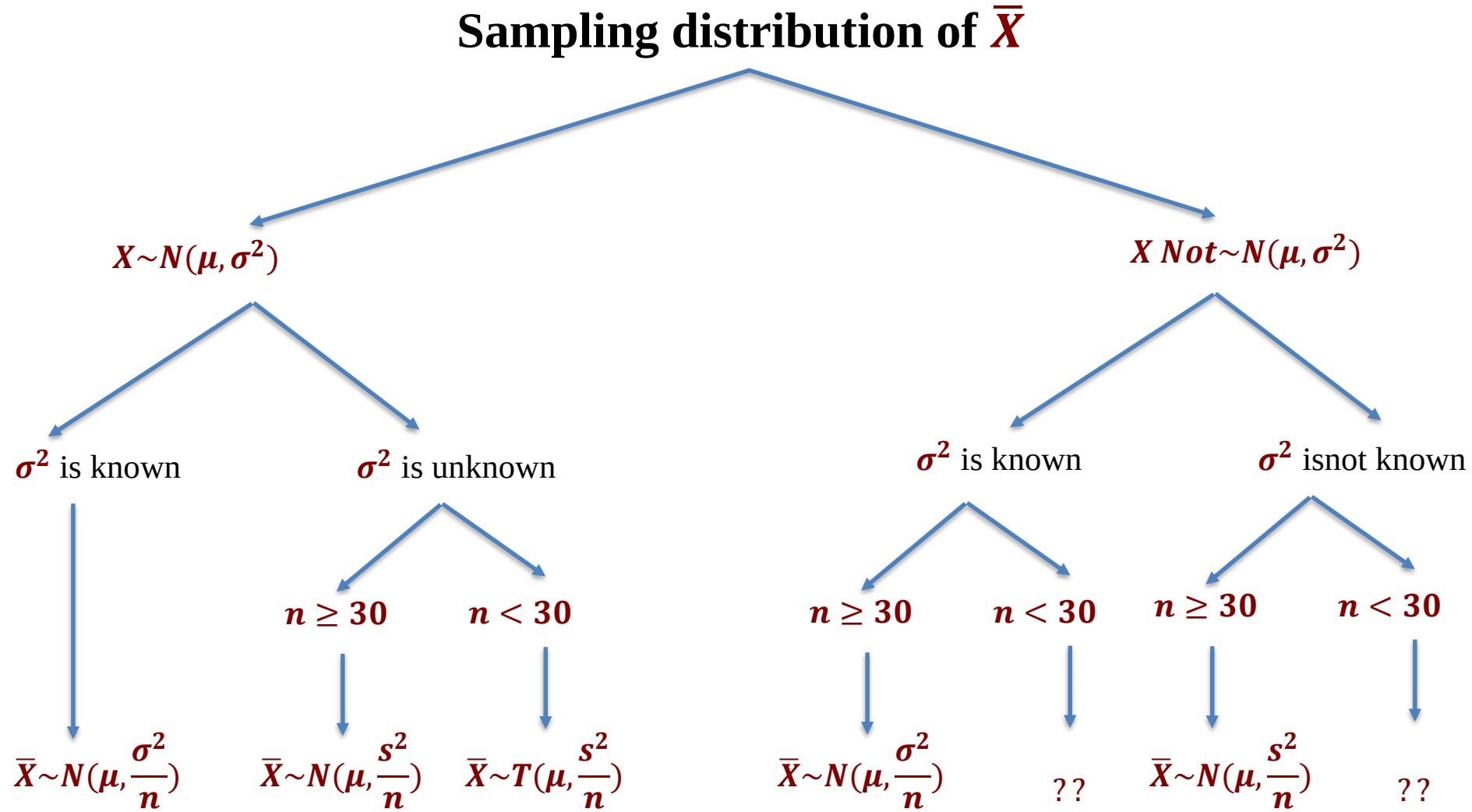
- For example $t_{0.05,6}$ is the t critical value that captures an upper-tail area of 0.05 under the *t* curve with 6 df .

- Because t curves are symmetric about zero, $-t_{\alpha,\nu}$ captures lower-tail area α ; i.e. $t_{1-\alpha,\nu} = -t_{\alpha,\nu}$

$$t_{0.5,\nu} = ?$$

Sampling Distribution of $\bar{X}$

- **Summary:**



Sampling Distribution of $\overline{X}$

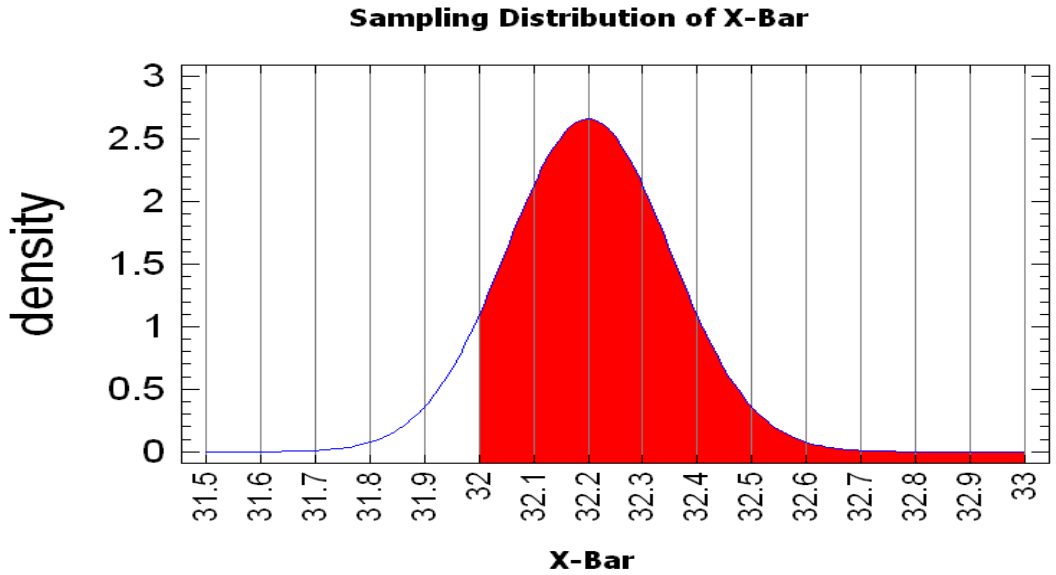
- The foreman of a bottling plant has observed that the amount of soda in each "32-ounce" bottle is actually a normally distributed random variable, with a mean of 32.2 ounces and a standard deviation of .3 ounce.
- If a customer buys a carton of **four** bottles, What is the distribution of the sample mean $\overline{X}$?
- what is the probability that the ***mean amount of the four bottles*** will be greater than 32 ounces?

Sampling Distribution of $\bar{X}$

$\bar{X}$ is normally distributed.

We want to find $P(\bar{X} > 32)$ where $\bar{X}$ is normally distributed $\mu = 32.2$ and standard error $= \frac{\sigma}{\sqrt{n}} = \frac{0.3}{\sqrt{4}} = 0.15$

$$P(\bar{X} > 32) = P\left(\frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{32 - 32.2}{.15}\right) = P(Z > -1.33) = .9082$$



Sampling Distribution of $\overline{X}$

Tests for an individual's intelligence quotient (IQ) are designed to be normally distributed, with a mean of 100 and a standard deviation of 15.

Let X denotes IQ test result

X is normally distributed with $\mu = 100$, $\sigma = 15$

a. What is the probability that an IQ result is more than 98?

P (X > 98) = ?

$$P(X > 98) = P\left(Z > \frac{98 - 100}{15}\right) = P(Z > -0.13) = 1 - P(Z < -0.13)$$
$$= 1 - 0.4483 = 0.5517$$

b. What is the probability that a randomly selected sample of 20 individuals would have a mean IQ of more than 105?

In this case, the probability needed is on the mean value:

$X \sim N(100, 15)$ $\qquad$ $\bar{X} \sim N(100, \frac{15}{\sqrt{20}})$

$$P(\bar{x} > 105) = P\left(Z > \frac{105-100}{15/\sqrt{20}}\right) = P(Z > 1.49) = 1 - P(Z < 1.49)$$

$$= 1 - 0.9319 = 0.0681$$

Sampling Distribution of $\overline{X}$

The mean selling price of a gallon of gasoline in the United States is \$1.30, with a standard deviation of \$0.28. What is the probability of selecting a sample of 36 gasoline stations and finding the sample mean within \$.08?

$$P(1.22 < \bar{X} < 1.38) = P\left(\frac{1.22-1.30}{\frac{0.28}{\sqrt{36}}} < Z < \frac{1.38-1.30}{\frac{0.28}{\sqrt{36}}}\right) = P(-1.71 < Z < 1.71) =$$

$$P(Z < 1.71) - P(Z < -1.71) = 0.9564 - 0.0436 = 0.9128$$

Sampling Distribution of $\bar{X}$

Knowing that, the student's grade follows the normal distribution with a mean of 80 and a standard deviation of 12. A simple random sample of 36 students was selected at random. Find the following probabilities:

i. The average grade is greater than 84?

$$P(\bar{X} > 84) = P\left(Z > \frac{84-80}{\frac{12}{\sqrt{36}}}\right) = P(Z > 2) = 1 - P(Z < 2) = 1 - 0.9772$$

ii. The average grade lies between 82 and 84?

$$P(82 < \bar{X} < 84) = P\left(\frac{82-80}{\frac{12}{\sqrt{36}}} < Z < \frac{84-80}{\frac{12}{\sqrt{36}}}\right) = P(1 < Z < 2) = P(Z < 2) -$$

$$P(Z < 1) = 0.9772 - 0.8413 = 0.1359$$

iii. The average grade is less than 84?

$$P(\bar{X} < 84) = P\left(Z < \frac{84-80}{\frac{12}{\sqrt{36}}}\right) = P(Z < 2) = 0.9772$$

Proportions

- **Definition 1:**

  - The population and sample proportions, denoted by $\boldsymbol{p}$ and $\boldsymbol{\hat{P}}$ respectively, are calculated as:

$$P = \frac{No.\ of\ elements\ in\ the\ population\ with\ specific\ characteristic}{Total\ number\ of\ elements\ in\ the\ population}$$

$$\hat{P} = \frac{X}{n} = \frac{No.\ of\ elements\ in\ the\ sample\ with\ specific\ characteristic}{Total\ number\ of\ elements\ in\ the\ sample}$$

  - The mean: $\mu_{\hat{P}} = P$, variance $\sigma^2_{\hat{P}} = \frac{P(1-P)}{n}$ **Why?**

  - According to the central limit theorem, the sampling distribution of $\hat{P}$ is approximately normal for a sufficiently large sample size ($Z = \frac{\hat{P}-P}{\sqrt{\frac{P(1-P)}{n}}} \sim N(0,1)$).

  - the sample size n is considered to be sufficiently large if nP and $n(1-P)$ are both greater than **5.**

Proportions

If it is known that the percentage of defective units is 5% in every shipment of some spare parts products. The QC inspector decided to reject a shipment of size 100 unit if the percentage of defective is greater than 8%. What is the probability of rejecting this shipment?

$$P(p > 0.08) = P\left(Z > \frac{0.08 - 0.05}{\sqrt{\frac{0.05 * (1 - 0.05)}{100}}}\right) = P(Z > 1.38) = 1 - P(Z < 1.38)$$

$$= 1 - 0.9162 = 0.0838$$

Proportions

A study was conducted by the National Health Club revealed that 30% of its new members are significantly overweight.

1. What is the probability that 175 or more of 500 new members are overweight?

$$P\left(p \geq \frac{175}{500}\right) = P(p \geq 0.35) = P\left(Z \geq \frac{0.35 - 0.3}{\sqrt{\frac{0.3 * (1 - 0.3)}{500}}}\right) = P(Z \geq 2.44)$$

$$= 1 - P(Z < 2.44) = 1 - 0.9927 = 0.0073$$

2. What is the probability that 140 or less of 500 new members are overweight?

$$\left(p \leq \frac{140}{500}\right) = P(p \leq 0.28) = P\left(Z \leq \frac{0.28 - 0.3}{\sqrt{\frac{0.3 * (1 - 0.3)}{500}}}\right) = P(Z \leq -0.98)$$

$$= 0.1635$$

Proportions

A survey was conducted by the Ministry of Health revealed that the proportion of all smoker college students is 30%. What is the probability that more than 40 students are smokers in a sample of 100 college students?

$$P(p > 0.4) = P\left(Z > \frac{0.4 - 0.3}{\sqrt{\frac{0.3 * (1 - 0.3)}{100}}}\right) = P(Z > 2.18) = 1 - P(Z < 2.18)$$

$$= 1 - 0.9857 = 0.0146$$

Point Estimation

A single value of a statistic that is used to approximate a population parameter is called a point estimate.

For example, an estimate of the population mean $\boldsymbol{\mu}$ denoted by $\hat{\boldsymbol{\mu}}$ is the sample mean $\overline{\boldsymbol{X}}$; i.e. $\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{X}}$.

Similarly; $\boldsymbol{S^2}$ is an estimate of $\boldsymbol{\sigma^2}$, i.e. $\hat{\boldsymbol{\sigma}}^2 = \boldsymbol{S^2}$

The statistic that one uses to obtain a point estimate is called an estimator and the value of the statistic is the estimate. An estimator is not expected to estimate the population parameter without error.

Interval Estimation

In many situations, a point estimate does not provide enough information about the parameter of interest.

It would be more desirable to determine an interval within which we would expect to find the value of the parameter. An interval of the form

$$L \leq \mu \leq U$$

Such an interval is called an **interval estimate**. The end points of this interval will be random quantities, since they are functions of the sample data.

Interval Estimation

In general, to construct an interval estimator of the unknown parameter $\boldsymbol{\theta}$, we must find two statistics $\boldsymbol{L}$ and $\boldsymbol{U}$ such that:

$$\boldsymbol{P(L \leq \theta \leq U) = 1 - \alpha}$$

where $\boldsymbol{0 < \alpha < 1}$.

$\boldsymbol{L}$ and $\boldsymbol{U}$ are called ***lower*** and ***upper confidence limits***, respectively, and the probability, **1- $\boldsymbol{\alpha}$,** is called the ***confidence coefficient*** (or the ***confidence level***).

The resulting interval $\boldsymbol{L \leq \theta \leq U}$ is called a $\boldsymbol{100(1 - \alpha)\%}$ **confidence interval** for $\boldsymbol{\theta}$.

i.e.; $\boldsymbol{1 - \alpha = 0.95}$, the interval is called $\boldsymbol{95\%}$ confidence interval for $\boldsymbol{\theta}$.

Interval Estimation

case (1): Confidence Interval for μ (the population is normal with known σ^2, or n is large)

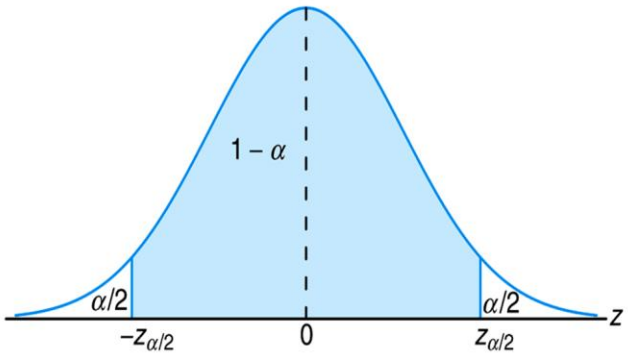
Let $X_1, X_2, \dots, X_n$ be a random sample of size **n** taken from a normal population with mean **μ** and variance σ^2**,** (or from non-normal population but **n** is sufficiently large) , we can establish a confidence interval for **μ** by considering the sampling distribution of $\bar{X}$ as follows:

Given : variance σ^2 of the population; n of samples; $\bar{X}$ sample mean; and α.

Required : ***lower (L)*** and ***upper confidence limits (U)***

Steps:

1. Find $Z_{\frac{\alpha}{2}}$; $\ni P\left(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$



Interval Estimation

2. $-Z_{\frac{\alpha}{2}} \leq \left(Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \right) \leq Z_{\frac{\alpha}{2}} \rightarrow \bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$

$$L = \bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}} \text{ and } U = \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$$

Where: $Z_{\frac{\alpha}{2}}$ is the z-value leaving an area of $\frac{\alpha}{2}$ to the right (i.e. $F\left(Z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$)

Example 4: A civil engineer is analyzing the compressive strength of concrete. Compressive strength is approximately normally distributed with standard deviation **32** psi. A random sample of **16** specimens has a mean compressive strength of **3250** psi. Construct a **95%** and **99%** confidence intervals on mean compressive strength. Compare between the width of the two intervals.

Interval Estimation

Solution:

To find **95%** C.I. for μ:

$$1 - \alpha = 0.95 \rightarrow \alpha = 0.05$$

Find $Z_{0.025} \ni F(Z_{0.025}) = 1 - 0.025 = 0.975 \rightarrow Z_{0.025} = 1.96$

Hence the **95%** confidence interval is:

$$3250 - 1.96 * \frac{32}{4} \leq \mu \leq 3250 + 1.96 * \frac{32}{4}$$

$$3234.32 \leq \mu \leq 3265.68$$

Similarly for **99%** confidence interval is:

$$3229.4 \leq \mu \leq 3270.6$$

We now see that the CI with confidence level **0.99** is wider than that with the **0.95** confidence level.

Estimation Example Mean (σ Known)

The mean of a random sample of $n = 25$ is $\bar{X} = 50$. Set up a 95% confidence interval estimate for μ if $\sigma = 10$.

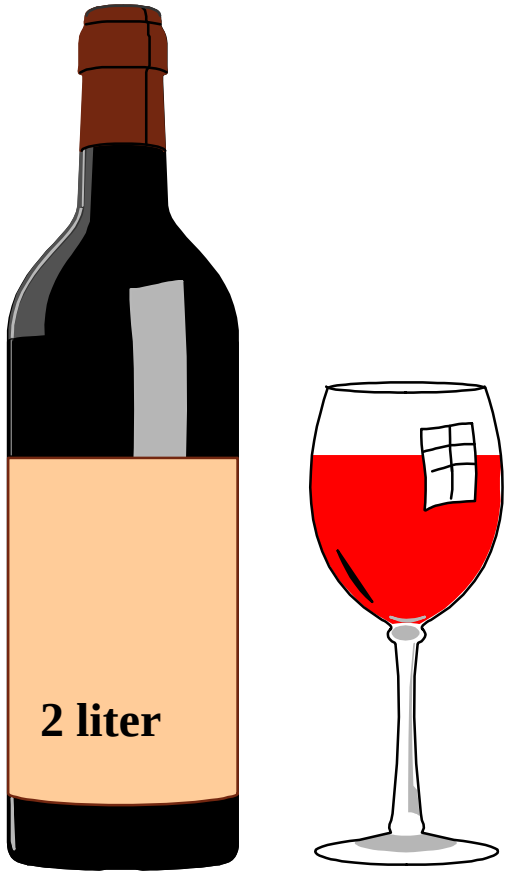
$$\bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$50 - 1.96 \cdot \frac{10}{\sqrt{25}} \le \mu \le 50 + 1.96 \cdot \frac{10}{\sqrt{25}}$$

$$46.08 \le \mu \le 53.92$$

Thinking Challenge

You're a Q/C inspector for Gallo. The σ for 2-liter bottles is **.05** liters. A random sample of **100** bottles showed $\overline{x}$ = **1.99** liters. What is the **90%** confidence interval estimate of the true **mean** amount in 2-liter bottles?



© 1984-1994 T/Maker Co.

Confidence Interval Solution*

$$\bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$1.99 - 1.645 \cdot \frac{.05}{\sqrt{100}} \leq \mu \leq 1.99 + 1.645 \cdot \frac{.05}{\sqrt{100}}$$

$$1.982 \leq \mu \leq 1.998$$

Example

- A computer company samples demand during lead time over 25 time periods:

235	374	309	499	253
421	361	514	462	369
394	439	348	344	330
261	374	302	466	535
386	316	296	332	334

- Its is known that the standard deviation of demand over lead time is 75 computers. We want to estimate the ***mean*** demand over lead time with 95% confidence in order to set

CALCULATE

- In order to use our confidence interval estimator, we need the following pieces of data:

$\bar{X}$	370.16
$Z_{\alpha/2}$	1.96
σ	75
n	25

Calculated from the data…

$1-\alpha = .95, \ \therefore \alpha/2 = .025$

$so \ \ z_{\alpha/2} = z_{.025} = 1.96$

Given

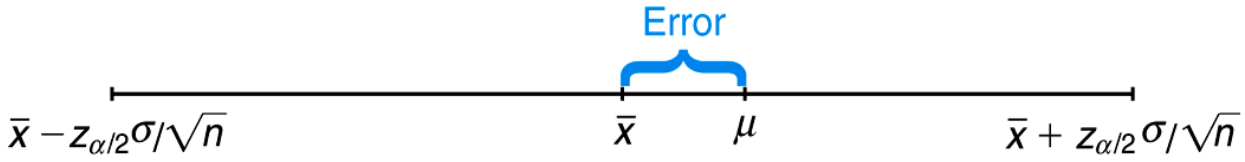
$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 370.16 \pm z_{.025}\frac{75}{\sqrt{25}} = 370.16 \pm 1.96\frac{75}{\sqrt{25}} = 370.16 \pm 29.40$$

Therefore:

- The **lower** and **upper** confidence limits are 340.76 and 399.56.

Interval Estimation

The $\mathbf{100(1-\alpha)\%}$ confidence interval provides an estimate of the accuracy of our point estimate. If $\boldsymbol{\mu}$ is actually the center value of the interval, then $\bar{X}$ estimates $\boldsymbol{\mu}$ without error. Most of the time, however, $\bar{X}$ will not be exactly equal to $\boldsymbol{\mu}$ and the point estimate is in error. The size of this error will be the absolute value of the difference between $\boldsymbol{\mu}$ and $\bar{X}$, and we can be $\mathbf{100(1-\alpha)\%}$ confident that this difference will not exceed $\frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$. We can readily see this if we draw a diagram of a hypothetical confidence interval



We are 100(1-α)% confident that the single computed interval $\bar{X} \pm \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$ contains the population mean μ

Interval Estimation

- ***Maximum Error and Sample size:***

$$P\left(-\frac{\sigma}{\sqrt{n}}Z_{\frac{\alpha}{2}} \leq \bar{X} - \mu \leq \frac{\sigma}{\sqrt{n}}Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(|\bar{X} - \mu| \leq \frac{\sigma}{\sqrt{n}}Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$|\bar{X} - \mu|$ is The error in estimating the mean μ

$E = \frac{\sigma}{\sqrt{n}}Z_{\frac{\alpha}{2}}$ is called the ***maximum error*** occurred in estimating the population mean μ with probability $(1 - \alpha)$.

Interval Estimation

- ***Maximum Error and Sample size:***

The formula for E can also be used to determine the sample size that is needed to attain a desired degree of precision. Solving the above equation of E for n we obtain

$$n = \left(\frac{\sigma Z_{\frac{\alpha}{2}}}{E} \right)^2$$

Example 5: The life, in hours, of a 150-watt light bulb is known to be approximately normally distributed with standard deviation 25 hours. What sample size should be taken in order to be 95% confident that the error in estimating the mean life is less than 5 hours.

Interval Estimation

Solution:

$\sigma = 25\ hours\ ; 1 - \alpha = 0.95\ ; E = 5\ hours$

$n =?$

$F\left(Z_{\frac{\alpha}{2}}\right) = 0.975 \rightarrow Z_{\frac{\alpha}{2}} = 1.96$

$$n = \left(\frac{\sigma Z_{\frac{\alpha}{2}}}{E}\right)^2 = 96$$

Example 6: A medical research worker intends to use the mean of a random sample of size 120 to estimate the mean blood pressure of women in their fifties. If, based on experience, he knows that σ =10.5 mm of mercury, what can he assert with probability 0.99 about the maximum error?

Interval Estimation

Solution:

$$\sigma = 10.5\ mm\ ;\ 1 - \alpha = 0.99\ ;\ n = 120$$

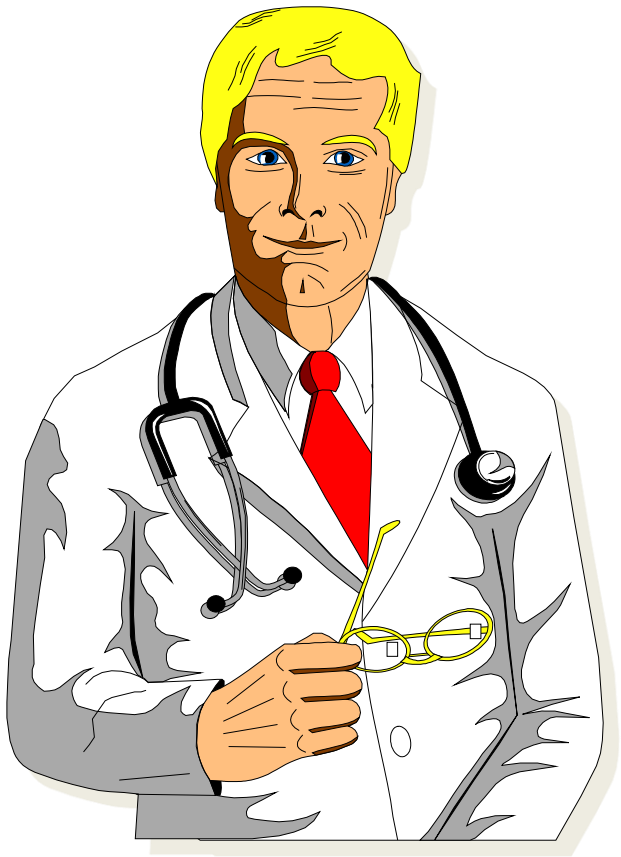
$$E = ?$$

$$F\left(Z_{\frac{\alpha}{2}}\right) = 0.995 \rightarrow Z_{\frac{\alpha}{2}} = 2.58$$

$$E = \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}} = 2.473\ mm$$

Thinking Challenge

You work in Human Resources at Merrill Lynch. You plan to survey employees to find their average medical expenses. You want to be **95%** confident that the sample **mean** is within **± $50**. A pilot study showed that σ was about **$400**. What **sample size** do you use?



Sample Size Solution*

$$n = \frac{(Z_{\alpha/2})^2 \sigma^2}{(SE)^2}$$

$$= \frac{(1.96)^2 (400)^2}{(50)^2}$$

$$= 245.86 \cong 246$$

Sample Size Example

What sample size is needed to be 90% confident the mean is within ± 5? A pilot study suggested that the standard deviation is 45.

$$n = \frac{(Z_{\alpha/2})^2 \sigma^2}{(SE)^2} = \frac{(1.645)^2 (45)^2}{(5)^2} = 219.2 \cong 220$$

Interval Estimation

- **Case (2) Confidence Interval for μ (n is small and the population is normal with unknown σ^2):**

we are attempting to estimate the mean of a population when the variance is unknown, and it is impossible to obtain a sample of size **n ≥ 30**. Cost can often be a factor that limits our sample size. As long as our population is approximately bell-shaped, confidence intervals can be computed by using the **t-distribution** in place of the standard normal distribution. The procedure is the same as for large samples except that we use values in place of **z** values and replace **σ** by its estimate **S**.

$$\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}} \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}$$

Interval Estimation

Example 7: Five measurements of the reaction time of an individual to certain stimuli were recorded as 0.28, 0.30, 0.27, 0.33 and 0.31 seconds. Find 99% confidence limits for the actual reaction time.

Solution:

$$\bar{X} = 0.298;\ S = 0.0239\ ;1 - \alpha = 0.99\ ;n = 5$$

$$L = ?\ ;U = ?$$

$$F\left(t_{\frac{\alpha}{2}}\right) = 0.995 \rightarrow t_{\frac{\alpha}{2}} = 4.6$$

$$L = \bar{X} - \frac{\sigma}{\sqrt{n}} t_{\frac{\alpha}{2}} = 0.249\ ;U = \bar{X} + \frac{\sigma}{\sqrt{n}} t_{\frac{\alpha}{2}} = 0.347$$

$$0.249 \leq \mu \leq 0.347$$

Estimation Example
Mean (σ Unknown)

A random sample of $n = 25$ has $\bar{x} = 50$ and $s = 8$.
Set up a 95% confidence interval estimate for μ.

$$\bar{X} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

$$50 - 2.064 \cdot \frac{8}{\sqrt{25}} \le \mu \le 50 + 2.064 \cdot \frac{8}{\sqrt{25}}$$

$$46.69 \le \mu \le 53.30$$

Thinking Challenge

You're a time study analyst in manufacturing. You've recorded the following task times (min.):
3.6, 4.2, 4.0, 3.5, 3.8, 3.1.
What is the **90%** confidence interval estimate of the population **mean** task time?

Confidence Interval Solution*

- $\overline{x} = 3.7$
- $s = 3.8987$
- $n = 6, \quad \text{df} = n - 1 = 6 - 1 = 5$
- $t_{.05} = 2.015$

$$3.7 - 2.015 \cdot \frac{3.8987}{\sqrt{6}} \le \mu \le 3.7 + 2.015 \cdot \frac{3.8987}{\sqrt{6}}$$

$$.492 \le \mu \le 6.908$$

Confidence Interval For The Proportion P.

Suppose there is a population of interest, a particular **trait** is being studied, and each member of the population can be classed **as either having or failing to have the trait (e.g. success or failure). Confidence limits are to be found for the parameter *P***, the proportion of the population with the trait. For this purpose, we should draw a random sample from the population of interest, determine the proportion of objects in the sample with the trait, and use this sample proportion as a point estimate of the population proportion **p**. That is,

$$\hat{P} = \frac{X}{n} = \frac{No.\,of\ elements\ in\ the\ sample\ with\ specific\ characteristic}{Total\ number\ of\ elements\ in\ the\ sample}$$

Confidence Interval For The Proportion P.

Note that:

- X is a binomial random variable with mean nP and variance $nP(1-P)$.
- As the sample size n increases ($nP \geq 5$ and $n(1-P) \geq 5$), the sampling distribution of $\hat{P}$ approaches a normal distributed with mean and variance,

$$\mu_{\hat{P}} = p, \ \sigma^2_{\hat{P}} = \frac{p(1-p)}{n}$$

i.e.

$$Z = \frac{\hat{P} - P}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

- A $100(1-\alpha)\%$ confidence limits for P are given by

$$\hat{P} \pm \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \, Z_{\frac{\alpha}{2}}$$

Interval Estimation

Example 8: In a Gallup poll, **491** randomly selected adults were asked whether they are in favor of the death penalty for a person convicted of murder, and 65% of them said that they were in favor.

i. Find the point estimate of the percentage of adults who are in favor of this death penalty.

ii. Find a **95%** confidence interval estimate of the percentage of adults who are in favor of this death penalty.

Solution:

Here we have, $\mathbf{n = 491}$, so the sample proportion is $\widehat{P} = 0.65$

i. the point estimate of the percentage of adults who are in favor of this death penalty = **65%**.

ii. $L = 0.608\,; U = 0.692$. **Why?**

Estimation Example Proportion

A random sample of 400 graduates showed 32 went to graduate school. Set up a 95% confidence interval estimate for *p*.

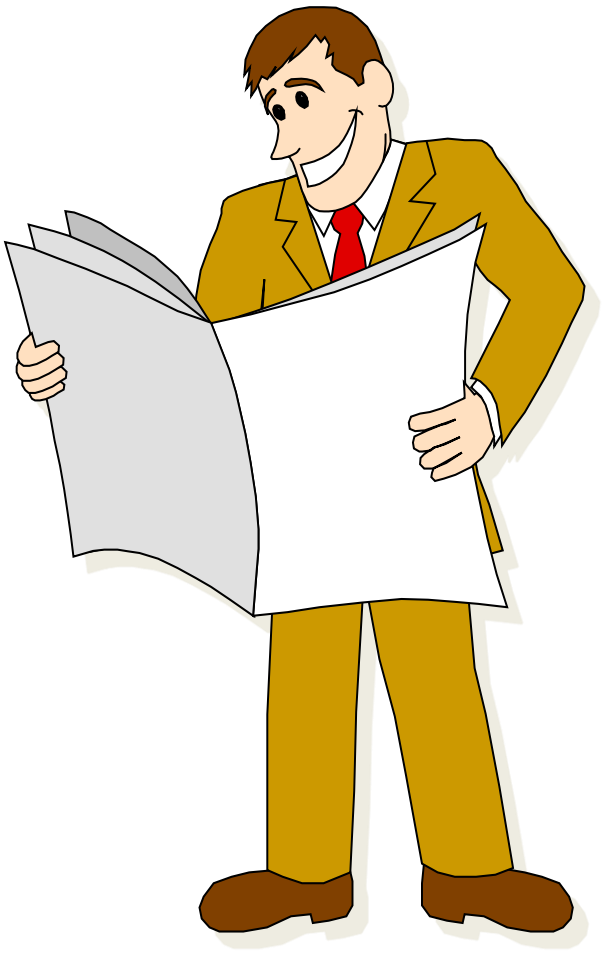
$$\hat{p} - Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \le p \le \hat{p} + Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$.08 - 1.96 \cdot \sqrt{\frac{.08 \cdot .92}{400}} \le p \le \quad .08 + 1.96 \cdot \sqrt{\frac{.08 \cdot .92}{400}}$$

$$.053 \le p \le .107$$

Thinking Challenge

You're a production manager for a newspaper. You want to find the % defective. Of **200** newspapers, **35** had defects. What is the **90%** confidence interval estimate of the population **proportion** defective?



Confidence Interval Solution*

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

$$.175 - 1.645 \cdot \sqrt{\frac{.175 \cdot (.825)}{200}} \leq p \leq \quad .175 + 1.645 \cdot \sqrt{\frac{.175 \cdot (.825)}{200}}$$

$$.1308 \leq p \leq .2192$$

Example

12

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. Therefore, estimate A 95% two-sided confidence interval for p ?

Solution a point estimate of the proportion of bearings in the population that exceeds the roughness specification is $\hat{P} = P_s = 10/85$

$$P_s - z_{0.025}\sqrt{\frac{P_s(1-P_s)}{n}} \le P \le P_s + z_{0.025}\sqrt{\frac{P_s(1-P_s)}{n}}$$

$$0.12 - 1.96\sqrt{\frac{0.12(0.88)}{85}} \le P \le 0.12 + 1.96\sqrt{\frac{0.12(0.88)}{85}}$$

$$0.05 \le P \le 0.19$$